4Г

УДК 004.272

С.Я. Гильгурт
ИПМЭ им. Г.Е. Пухова НАН Украины, г. Киев
hilgurt@ukr.net

Анализ применения унифицированных вычислителей в интеллектуальных системах

Рассмотрены существующие на сегодняшний день высокопроизводительные присоединенные вычислители, позволяющие при умеренных затратах существенно ускорить решение задач искусственного интеллекта.

Введение

При создании современных интеллектуальных систем, требующих высокопроизводительных расчетов, возникает проблема нехватки вычислительных ресурсов. Существующие суперкомпьютерные технологии остаются достаточно дорогостоящими решениями и не предоставляют возможности для действительно массового применения ресурсоемких интеллектуальных систем [1]. Получившие в последнее время широкое распространение вычислительные кластеры, несмотря на снижение стоимости и простоту реализации, тем не менее, не способны обеспечить требуемого уровня доступности и автономности.

В этой связи представляет повышенный интерес получившее в последнее время широкое распространение альтернативное направление развития вычислительных средств, связанное с повышением производительности компьютеров широкого применения как стационарных, так и мобильных, путем дооснащения вычислительными устройствами унифицированного, стандартизованного исполнения. В литературе появляется, особенно в последнее время, множество публикаций по всевозможным сопроцессорам, акселераторам, ускорителям вычислений, характеризующихся высокой степенью универсальности, то есть не ориентированных на решение узкого класса специализированных задач. Однако эти сведения разрознены и отрывочны, не позволяют в полной мере оценить текущее состояние в этой области.

В настоящей работе проанализирована информация по присоединенным вычислителям с целью оценки применимости для решения задач искусственного интеллекта.

Унифицированные вычислители

Использование специализированных вычислителей, или сопроцессоров, разработанных для решения узкого класса задач, является известным решением, давно применяющимся в технике, в том числе и для создания интеллектуальных систем. Однако узкая область применения, недостаточная гибкость, высокая стоимость разработки и сложности эксплуатации, связанные с нестандартностью таких устройств, существенно затрудняют их массовое использование.

В последние годы набирают популярность унифицированные вычислители (УВ), рассчитанные на решение относительно широкого круга задач и лишенные перечисленных выше недостатков узкоспециализированных устройств. Анализ имеющихся сведений позволяет выделить три наиболее перспективных направления в данной области в зависимости от типа УВ.

Блоки ускоренных вычислений APU

Первое направление базируется на применении единого изделия (в идеальном случае – в виде СБИС), которое содержит в себе несколько различных вычислительных архитектур, оптимизированных под ряд наиболее распространенных задач. Предполагается, что такой набор с большой степенью вероятности будет эффективен для большого числа вычислительных алгоритмов. Избыточность и простои незадействованных ресурсов предполагается компенсировать низкой стоимостью таких изделий, их широкой распространенностью и стандартизацией программного обеспечения.

Типичными представителями данного подхода являются так называемые блоки ускоренных вычислений APU (Accelerated Processing Units). Примером их практической реализации могут служить математические акселераторы семейства CSX600 Advance Board фирмы ClearSpeed Technology. Такого рода устройства широко обсуждались несколько лет назад, в настоящее время интерес к ним несколько снизился.

Основным недостатком унифицированных вычислителей класса APU является относительно невысокая гибкость и высокая стоимость. Этот подход, к сожалению, не дает принципиального разрешения противоречия между высоким быстродействием и универсальностью. Вопросы стандартизации и доступности массовому пользователю также остаются открытыми. Как следствие, интерес к подобным разработкам в последнее время заметно снизился.

Графические процессорные устройства GPGPU

Второе направление ориентировано на нестандартное использование в качестве математического ускорителя широкого применения какого-либо существующего компонента вычислительной системы. Главная ставка в этом случае делается на дешевизну и распространенность задействованного компонента, что существенно снижает стоимость владения техническим решением в целом. Классическим примером такого подхода является проведение ресурсоемких расчетов на обычных видеокартах ПЭВМ.

В настоящее время в вычислительном мире наблюдается настоящий бум, связанный с использованием для ресурсоемких вычислений графических адаптеров персональных компьютеров. Данное движение называют GPGPU (General-Purpose Computation on Graphics Processing Units) или ГПУ (графические процессорные устройства). Создаются стандарты, разрабатывается системное и прикладное программное обеспечение, исследуются вопросы распараллеливания вычислительных методов и алгоритмов. Производители аппаратуры начинают выпуск изделий, изначально ориентированных на поддержку технологии GPGPU. Примером такого устройства может служить анонсированный в 2008 году потоковый процессор FireStream 9250 фирмы AMD, содержащий 1 Гб оперативной памяти, пиковая производительность которого заявляется равной 1 Тфлопс при одинарной точности вычислений. Номинальная потребляемая мощность изделия не более 150 Вт.

К сожалению, УВ второго класса, в частности, GPGPU также свойственны недостатки, присущие всем вычислительным устройствам с жесткой архитектурой. Фиксированный набор инструкций, однозначно заданная связь между процессорными

элементами ограничивают гибкость и область возможного применения. Высокое энергопотребление и тепловыделение уже сейчас является серьезной проблемой, которая будет лишь усугубляться по мере дальнейшего наращивания производительности устройств GPGPU, затрудняя их интеграцию в системный блок универсального компьютера, тем более – ноутбука. Тем не менее, широкое распространение используемых компонентов и относительно хорошая проработанность вопросов стандартизации делают данный подход наиболее перспективным на ближайшие несколько лет. Главным же положительным эффектом от применения ГПУ представляются такие явления, как привлечение внимания большого числа разработчиков и пользователей к вопросам применения унифицированных вычислителей, постановка и решение многих вопросов по стандартизации и унификации и ряд других полезных фактов и тенденций.

РУВ на базе ПЛИС

Третье направление предполагает использование построенных на базе программируемых логических интегральных схем (ПЛИС) со статической памятью реконфигурируемых унифицированных вычислителей (РУВ), аналогом которых с определенными оговорками можно считать RA (Reconfigurable Accelerators) на базе FPGA.

Реконфигурируемые вычислители позволяют задействовать мощный потенциал программируемой логики. Современные ПЛИС содержат миллионы эквивалентных логических элементов и позволяют синтезировать в себе произвольную цифровую вычислительную структуру, способную функционировать на частотах в сотни и тысячи мегагерц [2], [3]. При таком подходе, теоретически, имеется возможность создавать вычислительное средство, идеально подходящее для каждой решаемой задачи.

В работе [4] обоснованы структура и состав реконфигурируемых унифицированных вычислителей. В настоящее время различными фирмами-производителями выпускается большое число изделий, которые могут быть использованы в качестве РУВ [5], например, реконфигурируемые ускорители RCHTX фирмы Celoxica, BenONE-PCIe фирмы Nallatech и ADPe-XRC-4 фирмы Alpha Data.

Применительно к интеллектуальным системам особый интерес представляют реконфигурируемые вычислители для переносных компьютеров. В качестве примера такого изделия можно назвать, например, ExpressCard-карточку расширения PICO E-16 производства фирмы Pico Computing. Данное устройство включает в себя ПЛИС Virtex-5 LX50 фирмы XILINX, сверхоперативное ЗУ емкостью 32 Мб, контроллер шины PCI Express, 18 одинарных либо 9 дифференциальных высокоскоростных каналов передачи данных. Потребляемая мощность изделия не более 3,45 Вт.

Унифицированные вычислители данного класса позволяют в значительной степени избавиться от недостатков упомянутых выше двух направлений. Главным сдерживающим фактором в распространении реконфигурируемых унифицированных вычислителей на базе программируемой логики следует признать необходимость трудоемкой процедуры создания загружаемых в ПЛИС конфигураций. Успешное решение этой проблемы, а также ряда вопросов по стандартизации конструктива и программного обеспечения различных уровней, позволят перейти к массовому производству и применению РУВ, что обеспечивает стабильно высокий интерес к данному направлению как одному из самых перспективных.

Интерфейсы обмена данными с УВ

В настоящей работе рассмотрены также вопросы взаимодействия УВ и традиционных процессоров. В табл. 1 приведены различные варианты обмена информацией между унифицированными вычислителями и остальными компонентами компьютера.

Таблица 1 – Интерфейсы обмена данными с УВ

№	Вариант обмена	Скорость, Мбайт/с
1	Как с периферийным устройством по шине USB	60
2	Как с накопителем данных по Serial ATA	до 75
3	По системной шине PCI / PCI-X / PCI Express	132 / 528 / 800
4	По шине доступа к памяти DDR-II PC2-6400	6400
5	По шине HyperTransport 2.0	41600

Каждый из вариантов обладает своими преимуществами и недостатками. Варианты расположены в порядке увеличения скорости обмена и соответственно стоимости решения. Рассмотрим их подробнее.

Интерфейсная периферийная шина USB позволяет задействовать большое число УВ, удобно в подключении к ПЭВМ (в том числе – в режиме «на лету»), дает возможность построения высокопроизводительной системы на базе мобильных компьютеров. К сожалению, данный вариант является самым медленным из рассматриваемых решений.

Подключение УВ подобно накопителям данных, в частности, через интерфейс Serial ATA, также не очень эффективно по быстродействию.

Использование для обмена с РУВ системных шин в большинстве случаев является оптимальным по соотношению производительность/стоимость решения. Число используемых вычислителей может быть достаточным для построения относительно больших реконфигурируемых систем. Этот вариант также является одним из самых надежных в плане защиты от морального старения, поскольку стандарты системных шин в современных универсальных ЭВМ являются одними из самых консервативных.

Вариант применения РУВ, конструктивно выполненных в виде модуля памяти на первый взгляд представляется привлекательным из-за высокого быстродействия. Однако ему свойственны существенные недостатки. Прежде всего, это небольшое количество свободных слотов, которые необходимо делить с модулями памяти. Кроме того, на размеры РУВ в этом варианте накладываются существенные ограничения.

Самым новым вариантом среди упомянутых является решение, основанное на технологии HyperTransport. По этой причине трудно провести объективную оценку данного варианта. Шина передачи данных HyperTransport изначально разрабатывалась для связи между собой процессоров, устройств памяти и реконфигурируемых вычислителей [6]. Из известных на сегодняшний день преимуществ можно назвать высокое быстродействие, а также возможность реализации УВ в конструктиве, аналогичном центральному процессору, то есть в виде модуля, вставляемого в процессорный сокет на материнской плате, изготовленной, например, в двухпроцессорном варианте. Примечательно, что в этом случае охлаждение вычислителя может осуществляться стандартным кулером, ориентированным на применение с ЦП. Очевидными недостат-

ками решения являются высокая стоимость и ограничение по числу одновременно подключенных УВ. Негативным также следует признать тот факт, что данная технология в настоящее время не поддерживается основным производителем микропроцессоров и наборов системной логики для персональных ЭВМ – фирмой Intel. Впрочем, данная компания также ведет работы в данном направлении.

Резюмируя вышесказанное, можно сделать вывод о том, что при отсутствии специфических требований оптимальным решением по соотношению «цена/производительность» на сегодняшний день следует считать обмен по системной шине PCI / PCI-X / PCI Express.

Выводы

В настоящей работе проанализированы существующие на сегодняшний день унифицированные вычислители, способные существенно ускорить решение интеллектуальных задач на стандартных ПЭВМ и ноутбуках. Выделено три направления, представляющих наибольший интерес для этих целей, проведено предварительное сравнение, рассмотрены преимущества и недостатки каждого из исследованных подходов. Наиболее перспективным для решения задач искусственного интеллекта представляется направление, ориентированное на применение РУВ. Рассмотрены также вопросы взаимодействия УВ с остальными компонентами универсальных компьютеров.

Литература

1. Воеводин В.В., Воеводин Вл.В. Параллельные вычисления. – СПб.: БХВ-Петербург, 2002. – 608 с.
2. Грушвицкий Р.И., Мурсаев А.Х., Угрюмов Е.П. Проектирование систем на микросхемах программируемой логики. – СПб.: БХВ-Петербург, 2002. – 608 с.
3. Палагин А.В., Опанасенко В.Н. Реконфигурируемые вычислительные системы: Основы и приложения – К.: Просвіта, 2006. – 280 с.
4. Гильгурт С.Я. Применение типовых устройств на базе программируемой логики для решения вычислительных задач // Труды II Междунар. конф. «Параллельные вычисления и задачи управления». – 2004. – М.: Институт проблем управления им. В.А. Трапезникова РАН. – 2004. – С. 514-530.
5. FPGA Boards and Systems. – Режим доступа: http://www.fpga-faq.com/FPGA_Boards.shtml
6. Гильгурт С.Я. О применении реконфигурируемых унифицированных вычислителей для решения научно-технических задач // Труды Междунар. научной конф. «Параллельные вычислительные технологии» (ПаВТ'2008). – 2008. – Челябинск: Изд. ЮУрГУ, 2008. – С. 358-363.

С.Я. Гільгурт
Аналіз застосування уніфікованих обчислювачів в інтелектуальних системах
Розглянуті існуючі на сьогоднішній день високопродуктивні приєднані обчислювачі, що дозволяють при помірних витратах істотно прискорити рішення задач штучного інтелекту.

S.Ya. Gyl’gurt
Analysis of the Utilization of Unificated Coprocessors in the Intellectual Systems
The families of modern high-performance coprocessors are investigated. Such technologies as APU, GPGPU and RC (URC – FPGA-based Unificated Reconfigurable Coprocessor) allow to raise greatly the productivity of artificial intelligence tasks solving with low costs.

Статья поступила в редакцию 30.09.2008.